Correction for Ambiguous Solutions in Factor Analysis Using a Penalized Least Squares Objective

Arkadiusz Sitek*, *Associate Member, IEEE*, Grant T. Gullberg, *Senior Member, IEEE*, and Ronald H. Huesman, *Senior Member, IEEE*

***Abstract*—Factor analysis is a powerful tool used for the analysis of dynamic studies. One of the major drawbacks of factor analysis of dynamic structures (FADS) is that the solution is not mathematically unique when only nonnegativity constraints are used to determine *factors* and *factor coefficients.* In this paper, a method to correct for ambiguous FADS solutions has been developed. A nonambiguous solution (to within certain scaling factors) is obtained by constructing and minimizing a new objective function. The most common objective function consists of a least squares term that when minimized with nonnegativity constraints, forces agreement between the applied factor model and the measured data. In our method, this objective function is modified by adding a term that penalizes multiple components in the images of the factor coefficients. Due to nonuniqueness effects, these factor coefficients consist of more than one physiological component. The technique was tested on computer simulations, an experimental canine cardiac study using ^{99m}Tc-teboroxime, and a patient planar ^{99m}Tc-MAG$_3$ renal study. The results show that the technique works well in comparison to the truth in computer simulations and to region of interest (ROI) measurements in the experimental studies.**

***Index Terms*—Dynamic SPECT, factor analysis, penalized least squares.**

I. Introduction

FACTOR analysis of dynamic structures (FADS) [1], [2], which uses a factor model of the dynamic data, is a semi-automatic technique used for the extraction of time activity curves (TACs) from a series of dynamic images. The *mixture analysis* method [3]–[5] is another application of the factor model to dynamic data. In this method, pixel-wise TACs are approximated by using a linear combination of underlying sub-TACs. The main difference between FADS and mixture analysis is that FADS extracts physiological TACs. More precisely, the obtained curves are interpreted as TACs of a given physiological region and the corresponding factor coefficients define the geometry of that region. In the mixture analysis method, the set of generated sub-TACs does not necessarily correspond to the underlying physiology. That being the case the set is used for a better representation than the pixel wise representation of the dynamic data [3]. We will concentrate on the FADS methods in this paper. Specifically, we will address the problem of nonuniqueness of the solution, which is the major drawback of FADS.

In the factor model of dynamic data, it is assumed that activity in each pixel is a linear combination of factors $\mathbf{F}$ with the coefficients of the linear combination defined in matrix $\mathbf{C}$. Using this assumption, the dynamic data $\mathbf{A}$ can be written as

$$\mathbf{A} = \mathbf{C}\mathbf{F} + \epsilon \tag{1}$$

with ϵ being an error in $\mathbf{A}$. The size of $\mathbf{A}$ is $N \times M$, where N is the number of pixels in the image and M is the number of dynamic images. The matrix of factors $\mathbf{F}$ is $P \times M$ and the matrix of the factor coefficients $\mathbf{C}$ is $N \times P$, with P being the number of factors. Put simply, it is assumed that the image is built from structures that have the same temporal behaviors. In cardiac imaging, such structures are the myocardium, blood pools, and liver. In renal imaging, such structures are the kidney cortex, the background, and the ureter.

The FADS method can be used to obtain operator independent results that have advantages over region-of-interest (ROI) measurements, which are obtained when an operator specifies ROIs that correspond to different physiological areas in the image. TACs obtained from ROI measurements may be composites of activities from different overlapping components in the selected ROI. These are the major disadvantage of ROI measurements. On the other hand, the FADS method allows separation of partially overlapping regions that have different temporal behaviors, and thereby enabling the extraction of TACs that correspond to those regions.

The FADS procedure usually consists of an orthogonal analysis of the dynamic sequence followed by an *oblique rotation.* The oblique rotation imposes nonnegativity constraints on the extracted TACs (*factors*) and the extracted images of those factors (*factor coefficients*) [1], [2]. Although the oblique rotation with nonnegativity procedure yields reasonable results, they are not unique [6], and depending on the dynamic study under consideration, the achieved solution may be quite different from the truth.

Manuscript received January 24, 2001; revised February 5, 2001. This work was supported by the National Heart, Lung and Blood Institute of the US Department of Health and Human Services under Grant R01-HL50663 and Grant P01-HL25840 and by the Director, Office of Science, Office of Biological and Environmental Research, Medical Science Division of the U.S. Department of Energy under Contract DE-AC03-76SF00098. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was J. Fessler. *Astreisk indicates corresponding author.*

*A. Sitek was with the E. O. Lawrence Berkeley National Laboratory, One Cyclotron Rd. MS55-121, Berkeley, CA 94720 USA. He is presently with the Radiology Department, University of Utah, 729 Arapeen Dr., Salt Lake City, UT 84108-1218 USA, and the Radiology Department, Beth Israel Deaconess Medical Center, Rabb Bldg. L-345, 330 Brookline Ave., Boston, MA 02215 USA (e-mail: asitek@caregroup.harvard.edu).

G. T. Gullberg is with the Radiology Department, University of Utah, Salt Lake City, UT 84108-1218 USA.

R. H. Huesman is with the E. O. Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA.

Publisher Item Identifier S 0278-0062(02)04084-3.

0278-0062/02$17.00 © 2002 IEEE

To explain nonuniqueness of the factor model lets consider a data set with two factors. Using (1) the data are $A = C_1F_1 + C_2F_2$ where C_1 and C_2 are the factor coefficients for factors F_1 and F_2, respectively. The above equation can be rearranged to $A = (C_1 + aC_2)F_1 + C_2(F_2 - aF_1)$ with a being some constant. If only nonnegativity constraints are used then the factor coefficients $C_1 + aC_2$ and C_2 and factors F_1 and $F_2 - aF_1$ describe the same data set A as do the factor coefficients C_1 and C_2 and factors F_1 and F_2 as long as $C_1 + aC_2$ and $F_2 - aF_1$ are nonnegative.

As seen in the above example, it is easy to construct an additional set of factor coefficients and factors from the existing set, as long as conditions $C_1 + aC_2 \geq 0$ and $F_2 - aF_1 \geq 0$ are satisfied. This simple example illustrates the nonuniqueness problem for two factors, but similar nonuniqueness considerations apply to a model with more factors. The severity of the nonuniqueness artifacts depends on the TACs of the physiological components; the nonuniqueness effects can be very serious in one study and nearly nonexistent in another. For a detailed mathematical analysis and more discussion on these effects, see [7]. Nonuniqueness is a very serious drawback to the FADS method.

To correct for nonuniqueness, additional *a priori* information about the data being analyzed can be used [7]–[12]. Information about the spatial distribution of the factor coefficients can be used as *a prior* information. For example, the user may be required to specify a region in the image (such as pure background or pure blood) where only one component is present, methods that use *a priori* information cannot be used generically, they are designed to work only for specific kinds of dynamic studies. Another way to correct for nonuniqueness is to use the maximum entropy principle [13], [14]. A different approach to correcting ambiguous solutions was considered in [7], wherein a postprocessing technique was applied to the solution of the nonunique FADS. However, this approach also uses *a priori* information about the spatial distribution of factors in cardiac imaging. This technique was applied to Tc-99m-teboroxime cardiac imaging. Another example of correcting for nonuniqueness can be found in mixture analysis in which unitary constraints [5] have been used to improve the estimation of the background component, thereby enabling us to resolve other components.

In this paper, we developed a factor analysis method that does not require orthogonal analysis. The method does not use any *a priori* information and is not restricted to a specific type of imaging. It can be applied to any medical dynamic sequence of images. The technique involves using the penalized least squares (PLS) objective function to uniquely (to within P scale factors) and accurately extract factors and factor coefficients.

There are three terms in the objective function. One term is the least squares (LS) term, which forces agreement between the acquired data and the applied factor model. The second term imposes the nonnegativity constraints on the factors and the images of factor coefficients as they are being acquired. The third term makes the result of the minimization of the objective function unique by minimizing the products between the factor coefficient images. The rationale behind the choice of the third term will be given in Section II. The method was tested on computer simulations. It was also used to analyze a canine ^{99m}Tc-teboroxime cardiac study and a patient ^{99m}Tc-MAG$_3$ renal study.

II. Methods

A. PLS-FADS

The LS objective, f_{LS}, is a Cartesian norm between the measured the data and the factor model described by

$$f_{\mathrm{LS}}\left(\hat{\mathbf{C}}, \hat{\mathbf{F}}\right) = \sum_{i=1}^{N}\sum_{t=1}^{M}\left(\sum_{p=1}^{P}\hat{C}_{ip}\hat{F}_{pt} - A_{it}\right)^2 \tag{2}$$

where $\hat{F}_{pt}$ is the estimate of the pth factor and t is an index in time. $\hat{C}_{ip}$ is the ith pixel of the estimate of the pth factor coefficient image. A_{it} represents the value of the measurement data (dynamic sequence) at the ith pixel in the tth time frame. If $\hat{\mathbf{C}}$ and $\hat{\mathbf{F}}$ are to be physiologically meaningful they must be nonnegative. To impose the nonnegativity of estimates $\hat{\mathbf{C}}$ and $\hat{\mathbf{F}}$, the LS objective is modified by the term $f_{\mathrm{neg}}(\hat{\mathbf{C}}, \hat{\mathbf{F}})$, defined as

$$f_{\mathrm{neg}}\left(\hat{\mathbf{C}}, \hat{\mathbf{F}}\right) = \sum_{i,p=1}^{N,P} H\left(\hat{C}_{ip}\right) + \sum_{t,p=1}^{K,P} H\left(\hat{F}_{pt}\right) \tag{3}$$

where

$$H(x) = \begin{cases} ax^2 & x < 0 \\ 0 & x \geq 0 \end{cases} \tag{4}$$

with a being a penalty constant. By minimizing $f_{\mathrm{LS}}(\hat{\mathbf{C}}, \hat{\mathbf{F}}) + f_{\mathrm{neg}}(\hat{\mathbf{C}}, \hat{\mathbf{F}})$, results similar to standard FADS with oblique rotation and nonnegativity constraints are obtained [7]. The LS method with nonnegativity constraints will be referred to as the LS-FADS method. Nonnegativity alone though is not enough to guarantee that each factor coefficient image corresponds to a single physiological region; by physiological region we mean the region in the image that has the same temporal behavior. For example, in cardiac imaging such physiological regions are the left ventricular (LV) blood pool, right ventricular (RV) blood pool, liver, and myocardial tissue (if there is no abnormal uptake in the heart muscle due to infarctions).

Images of the factor coefficients obtained using FADS should correspond to the images of different physiological regions. Ideally, it is expected that in each factor coefficient image only one physiological region will be present. However, due to the nonuniqueness effect, each of the obtained images can be a linear combination of physiological regions in the image. In other words, each coefficient image acquired using nonunique FADS may be a mixture of multiple true physiological components (for more details and derivation see [7]). In order to reduce the amount of mixing, we add to the objective function a term that is a dot product of the normalized coefficients

$$f_{\mathrm{uni}}(\hat{\mathbf{C}}) = b\sum_{p=1}^{P}\sum_{q=p+1}^{P}\sum_{i=1}^{N}\frac{\hat{C}_{ip}}{\sqrt{\sum_{j=1}^{N}\hat{C}_{jp}^2}}\cdot\frac{\hat{C}_{iq}}{\sqrt{\sum_{j=1}^{N}\hat{C}_{jq}^2}}. \tag{5}$$

A total objective function $f_{\mathrm{PLS}} = f_{\mathrm{LS}} + f_{\mathrm{neg}} + f_{\mathrm{uni}}$ will be called a penalized LS objective function. Minimization of

(5) will reduce the amount of overlap between different coefficient images and therefore will minimize the amount of mixing between the different components. It is important in f_{uni} that the values of the image coefficients be normalized in such a way that the scaling of $\hat{\mathbf{C}}$ does not affect the value of f_{uni}. In (5), the values are normalized by $\sqrt{\sum_{j=1}^{N} \hat{C}_{jp}^2}$ in the denominator. Without this normalization, f_{uni} could be scaled to zero by scaling $\hat{\mathbf{C}}$ to zero. Such scaling of $\hat{\mathbf{C}}$ is allowed in the factor model since there is a multiplicative relation between $\hat{\mathbf{C}}$ and $\hat{\mathbf{F}}$. Therefore, by scaling $\hat{\mathbf{C}}$ by a constant x, the matrix $\hat{\mathbf{F}}$ is scaled by $1/x$ and the model expressed by (1) holds. This scaling creates additional degrees of freedom which are not constrained by the objective function.

The coefficient images that result from the minimization of f_{PLS} should have simple structures that correspond to single physiological regions. The lower value of f_{uni}, compared with $f_{\text{LS}} + f_{\text{neg}}$, was chosen so that the nonnegativity of the results imposed by f_{neg}, as well as agreement with the data guaranteed by f_{LS}, are not compromised by the addition of the f_{uni} term to the objective function. The constant b is applied in (5) to adjust the strength of the nonuniqueness penalty. The optimal value of b was investigated in computer simulations for different levels of noise.

B. Numerical Optimization

First, a number of factors P had to be chosen before the minimization was performed. We used an orthogonal singular value decomposition of the dynamic data to examine magnitudes of the singular values. Then based on the number of singular values, which were well above the noise level, we chose the appropriate number of components (see also [14]).

We found from our experience that the algorithm is not sensitive to the selection of a starting point. For all optimization results included in this paper, the same starting point was used. We initialized all images of factor coefficients with a value of one. Since optimization was done by a gradient method, and the values of $\hat{\mathbf{C}}$ were initialized by a constant, the factors had to be initialized with linearly independent functions. Therefore, to avoid stalling the optimization, any row of $\hat{\mathbf{F}}$ could not be a linear combination of the other rows. Also the values of $\hat{\mathbf{F}}$ were chosen such that the resulting initial values of the matrix $\mathbf{A} = \hat{\mathbf{C}}\hat{\mathbf{F}}$ had approximately the same order of magnitude as the values in the matrix $\mathbf{A}$ of the dynamic data being analyzed.

After the initialization, a conjugate gradient method [15] was used for the minimization. In each iteration, a gradient of the objective function $\bigtriangledown f_{\text{PLS}} = (\partial f_{\text{PLS}}/\partial \hat{\mathbf{C}}, \partial f_{\text{PLS}}/\partial \hat{\mathbf{F}})$ was calculated analytically using Eqs. (2), (3), and (5). Then the function was minimized in the conjugate direction of the gradient using the Brent method [15]. The iterations were terminated when the relative change in the objective function in one iteration was less than 10^{-6}. Depending on the size of the study, 40–150 iterations were required. In terms of speed, the algorithm took approximately 20 s to converge on the SUN Ultra 1 system for the two-dimensional (2-D) data sets.

The method which we used for selecting b was to adjust the value of the penalty parameter after every few iterations of the conjugate gradient algorithm so that the ratio of $f_{\text{uni}}/(f_{\text{LS}} + f_{\text{neg}})$ was equal to 0.1. The value of a was was set to some large value. Its value was not critical when the algorithm was used to estimate physical (no negative values) solutions.

After the optimization, the results and matrices $\hat{\mathbf{C}}$ and $\hat{\mathbf{F}}$ were re-scaled. That is, the coefficients $\hat{\mathbf{C}}$ were scaled so that all values of the coefficients were in the range from zero to one. This scaling was done separately for each column in $\hat{\mathbf{C}}$ by finding the approximate maximum value of each column by averaging the ten pixels with the highest coefficients, then dividing all coefficients of a given column by this maximum. Obviously, the corresponding rows in $\hat{\mathbf{F}}$ need to be scaled by the reciprocal of this maximum in order for (1) to hold.

C. Computer Simulations

A simple dynamic sequence was created from the three objects presented in Fig. 1 (first row). The intensities of each object were changed according to the curves presented in Fig. 1(A)–(C). A total of 90, 64 × 64 pixel images were generated. Noise was not added to the images.

A more realistic computer simulation was also performed. A slice of the MCAT phantom [16] was used for this simulation. Three components in the image were simulated: The RV blood pool, the LV blood pool, and the myocardial tissue (TI). The presence of vascular activity in the myocardial tissue was simulated by adding 10% of LV activity to the tissue. The simulated curves are presented as truth in Fig. 2(A)–(C), for the RV, LV, and tissue, respectively. A previously developed model, and its parameters, were used to create the simulated curves [7]. Partial volume effects were simulated by smoothing the images of the MCAT phantom components so that the neighboring structures partially overlap by two to three pixels.

A dynamic sequence of 180, 20 × 20 pixel images was created and analyzed by LS FADS. Dynamic sequences with normally distributed noise (variances equal to 15%, 25%, and 35% of the value of the mean) were generated from a noise-free sequence. The selection of normally distributed noise for the simulation was based on the fact that the distribution of noise in reconstructed images is unknown, and we assumed that normally distributed noise gives a reasonable approximation. For each computer simulation with noise, 100 noise realizations were used.

The accuracy of the curve estimates was measured using a measure D, which was the total distance from the true LV and RV curves to the estimates of the LV and RV curves obtained using the LS FADS method. This measure was defined as

$$D = \sum_{p=\text{LV, RV}} \sum_{t=1}^{M} \frac{\left|\hat{F}_{pt} - F_{pt}\right|}{\sum_{p'=\text{LV, RV}} \sum_{t'=1}^{M} F_{p't'}} \tag{6}$$

where $\mathbf{F}$ is the true factor and $\hat{\mathbf{F}}$ is an estimate of that factor obtained via the LS FADS method. The measure D was calculated over the LV and RV only. The tissue component was not taken into account during the calculation of the measure D because the error calculated by D of the tissue curve occurs mainly as a result of the ambiguity over how much simulated vasculature in

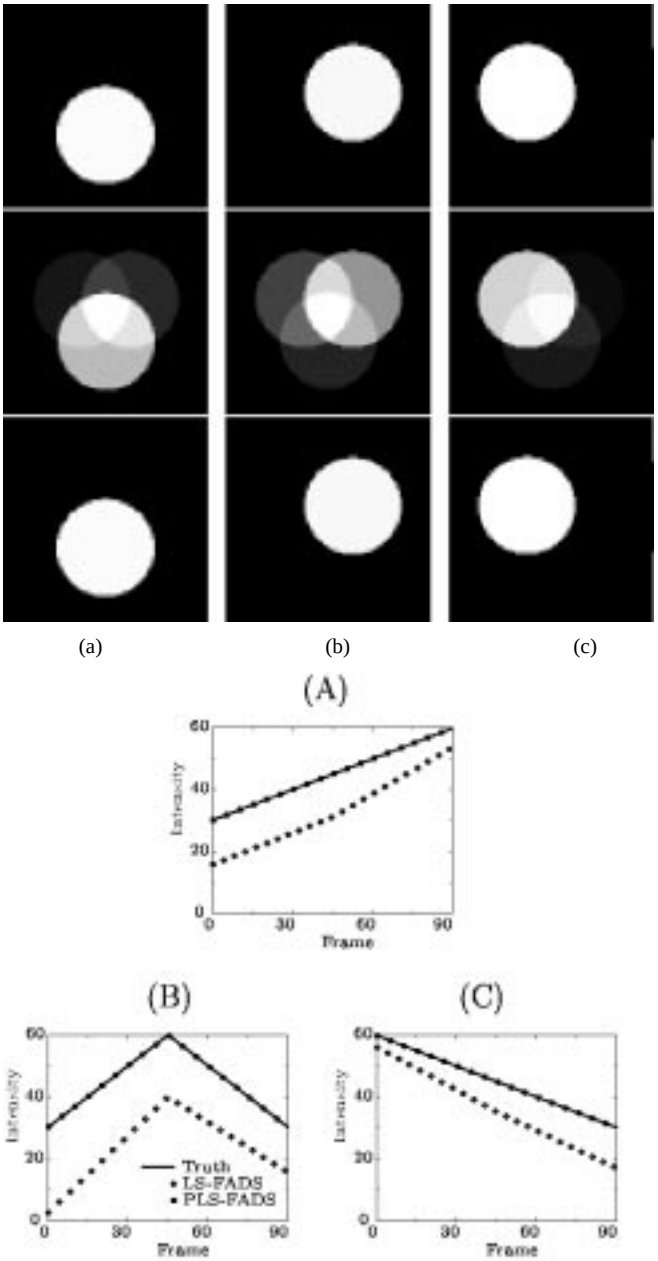

Fig. 1. Results of computer simulations. The top row of images corresponds to the circular objects which were used in the computer simulations. The second row present the factor coefficient images [these images correspond to columns in the matrix $\hat{\mathbf{C}}$ in (1)] obtained by the LS-FADS method. The last row presents the factor coefficients images obtained by the PLS-FADS method. The plots present a comparison of factors [rows in matrix $\hat{\mathbf{F}}$ in (1) obtained with different factor analysis methods to the true curves. Symbols correspond to every 4th frame. Plots (A), (B), and (C) correspond to objects presented in columns (a), (b), and (c), respectively.

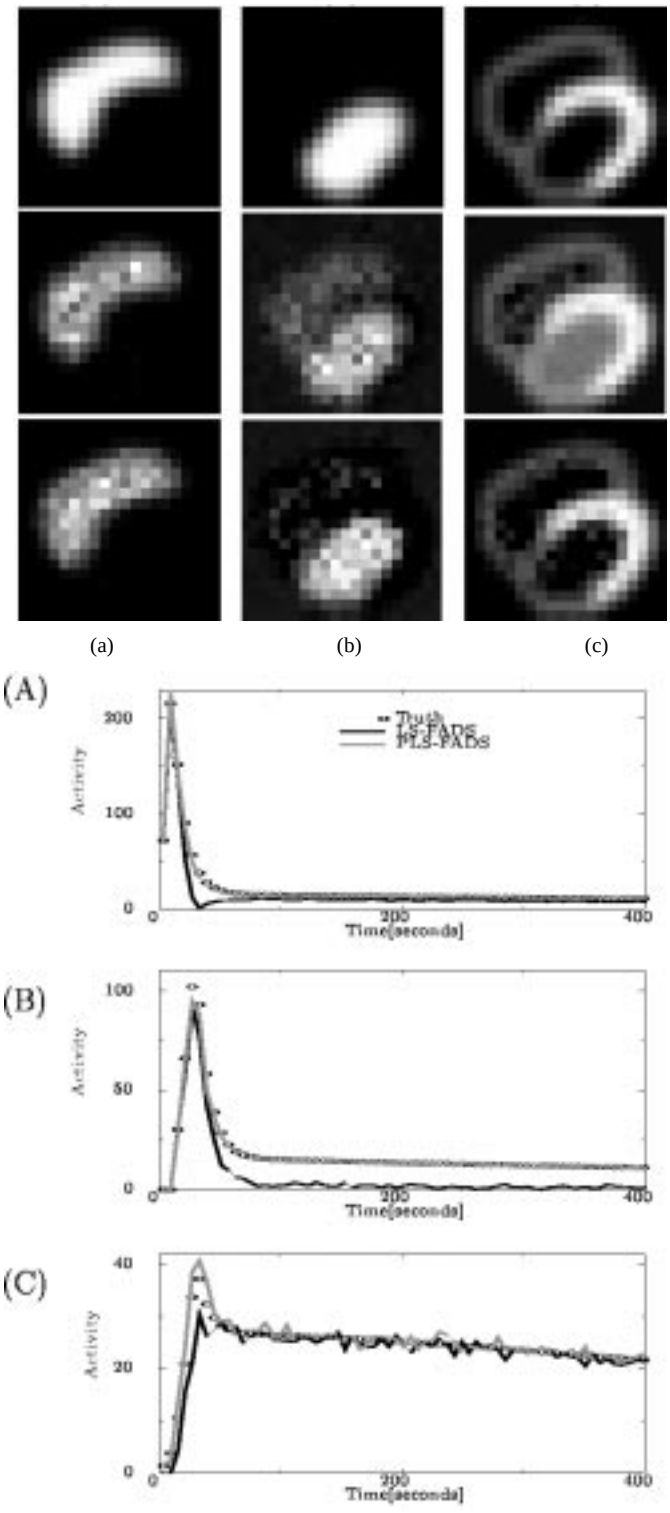

Fig. 2. Images of factor coefficients used in computer simulations (MCAT phantom). (a) RV, (b) LV, and (c) tissue. In the second and the third rows, images of factors coefficients obtained by LS-FADS and PLS-FADS methods are presented. Plots (A), (B), and (C) present the factors obtained by the LS-FADS and PLS-FADS methods with comparison to the true simulated curves for the RV, LV, and tissue, respectively. The same gray scale representing values from zero to one is used for each image. The presented results are for one realization of 25% noise. For the LS-FADS method the penalty parameter was equal to zero, and for the PLS-FADS method, b equaled to 2×10^5. Only the first 400 s out of a total 1080 s of the curves are presented.

the tissue is present in the tissue curve obtained by nonunique LS-FADS. Since this ambiguity is not directly connected to the nonuniqueness effects, when considering the LV and RV components alone, we only take into account the effects of nonuniqueness of the factor model.

D. Experimental Studies

Data from a ^{99m}Tc-teboroxime canine study were analyzed. A three-detector IRIX scanner (Marconi Imaging Systems, Inc., Cleveland, OH) was used to acquire transmission and emission projection data. The camera acquired 120 projections every 6

s for approximately 18 min. The 179 dynamic images were reconstructed using 25 iterations of the maximum-likelihood expectation-maximization algorithm with attenuation correction. The reconstructed three-dimensional (3-D) images were then reoriented to obtain short-axis slices of the heart. Factor analysis methods were applied to a 2-D, 11 $\times$ 11 pixel region that encompassed a short axis slice of the myocardium. A 3-D analysis in which a series of 179 (six slices, 11 $\times$ 11 pixels) images were analyzed as a 3-D data set was also performed.

Patient data from a planar ^{99m}Tc-MAG$_3$ renal study were also analyzed. The data were acquired using an eCam system (Siemens, Hoffman Estates, IL). The patient rested in a supine position as data were acquired by the detector in a 128 $\times$ 128 pixel matrix. The 300 dynamic images were acquired every 5 s. FADS was applied to an 18 $\times$ 20 pixel region that encompassed the right kidney. Only small regions of the image were investigated since incorporation of the larger regions would require increasing the number of factors to adequately represent the data, which would decrease the accuracy of the obtained factors that we are interested in.

ROI measurements were performed in both experimental studies. In the 2-D canine cardiac study, the ROIs spanned four pixels. They were automatically determined using FADS results that identified pixels with the highest values of factor coefficients (matrix $\mathbf{C}$) that corresponded to the LV, RV, and tissue. Such semi-automatic selection of ROIs decreases the amount of errors caused by overlapping of neighboring factors. It is also user independent. For the 3-D cardiac and renal patient study the ROIs spanned ten pixels, and the method for determining the locations of the ROIs was the same as in the 2-D cardiac study.

The change of contrast in the image of the factor coefficients between region one in the image and region two in the image was measured using the following definition:

$$\tilde{c}_2^1 = \frac{\mathbf{C}(1) - \mathbf{C}(2)}{\mathbf{C}(1)} \tag{7}$$

where $\mathbf{C}(1)$ and $\mathbf{C}(2)$ are average values of three pixels from a given region in the factor coefficient image $\mathbf{C}$ for regions one and two, respectively.

III. Results

The conjugate gradient algorithm was very robust. Unconstrained degrees of freedom due to scaling did not hinder convergence of the algorithm. All results are presented in the forms of images that correspond to images of factor coefficients and curves that correspond to factors. Since the results were rescaled after the reconstruction, as described in Section II-B, all images are in the range from zero to one, so the same gray scale is used for all of them.

A. Results of Computer Simulations

The results of the LS-FADS study are presented in Fig. 1. The images of factor coefficients obtained using this method are shown in second row of Fig. 1. In each image, all of the objects can be seen due to the nonuniqueness effects. The factors obtained using the LS-FADS method are presented in Fig. 1(A)–(C). They all show substantial disagreement with the simulated factors. The third row of images in Fig. 1 correspond to images of factor coefficients obtained using the PLS-FADS method. Factors obtained by both methods are presented in Fig. 1(A)–(C) and compared with simulated curves. Both, the images and the curves obtained using the PLS-FADS method show very strong agreement with the simulated objects.

The results of the more realistic simulation are presented in Fig. 2. The first row of images presents the factor coefficient images used to simulate teboroxime uptake in the heart. The second row of images corresponds to factor coefficient images obtained using the nonunique LS-FADS method. Nonuniqueness artifacts similar to those shown in previous computer simulations are clearly visible. For instance, in the image of the LV, some of the RV can be seen, and in the image of the tissue the LV is clearly visible. Application of the penalized objective reduces these artifacts and creates much better agreement between the factors obtained through the FADS methods and the true factors, as can be seen in the third row of Fig. 2(a)–(c).

The value of the error measure D is plotted versus the strength parameter b in Fig. 3(a). For the low values of b (less than 10^4) reconstructions yield larger errors (high values of D) because the nonuniqueness correction has little effect on the final results since the value of b is low. D slowly decreased when FADS was applied with $b \sim 5 \times 10^3$. Further increases of b to $\sim 5 \times 10^5$ make D rapidly increase because the domination of the term f_{uni} in the objective function. High b forces the dot product between the images of the different factor coefficients to be close to zero. The null value of the dot product term in the high b results is achieved by creating sharp edges between components, i.e., the pixels that normally belong to two neighboring components, due to the partial volume effect, are forced to be in one or the other of the neighboring structures. The rapid degradation of the results, seen in Fig. 3 as a sharp increase of D, is created by further increases of b, which force a negative value on the f_{uni}. Negative values of f_{uni} can be achieved when values of the pixels in one of the images of the factor coefficients reaches slightly negative values. As a result, the nonnegativity term is not increased significantly, and at the same time the dot product of this image with other nonnegative components causes a negative contribution to f_{uni}.

In Fig. 3(a), for some values of b the standard deviation of the calculated D is high and the distribution is asymmetric. This finding is illustrated in Fig. 3(b). In the histogram, it can be seen that the final value of D for different noise realizations for one value of b is either high or low. This makes the distribution of D high and asymmetric. Fig. 3(c) presents a comparison of the $D(b)$ relationship for different noise levels. It shows that with higher noise the best achieved D is higher and the range of b, for which the nonuniqueness correction works, is narrower.

Table I presents the summary of the computer simulation results. It shows that the use of the nonuniqueness penalty greatly improves the value of the measure D. In Table I, values of the penalty parameter, b and $f_{\text{uni}}/f_{\text{LS}}$ are given for the PLS-FADS method, which derived the best value of D. The table also shows

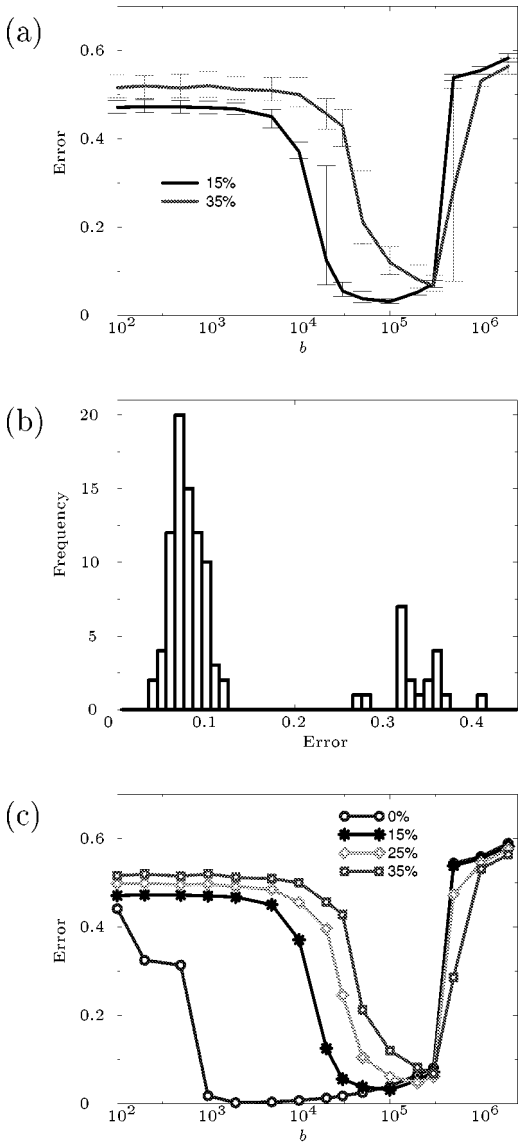

Fig. 3. The Error, D, of the reconstruction of the RV and LV blood curves versus the value of the penalty parameter b [graph (a) and (c)]. The error bars in (a) correspond to values of the standard deviation estimated using 100 noise realizations. In (c), four curves are presented which correspond to four different levels of simulated normal noise with variances equal to 0%, 15%, 25%, and 35%. The histogram in (b) presents the values of the Error D for $b = 2 \times 10^4$ at a noise level of 15%, for which the distribution was highly asymmetric [see (a)].

that the ratio of $f_{\text{uni}}/f_{\text{LS}}$ remains at approximately the same level, 10%, even though the noise levels change considerably.

B. Results of Experimental Studies

The results of the 2-D cardiac canine study are summarized in Fig. 4. The images in Fig. 4 represent factor coefficients for three different factors corresponding physiologically to the RV, LV, and the myocardial tissue. The first row displays the results obtained using the LS-FADS method. The second row displays the PLS-FADS results. The contrast is improved in the PLS-FADS images of the LV and tissue (contrast between pixels corresponding to the LV and tissue) in comparison to the images obtained using the LS-FADS method. For the LV coefficient image, $\tilde{c}^{\text{LV}}_{tissue}$ was 0.79 for the LS-FADS and 0.95 for the PLS-FADS. The value of $\tilde{c}^{\text{LV}}_{\text{RV}}$ also improved from 0.65 for the LS-FADS method to 0.86 for the PLS-FADS method. In the image of tissue coefficients, $\tilde{c}^{\text{tissue}}_{\text{LV}}$ changed from 0.84 to 1.00. Fig. 4(A)–(C) show factors obtained using the LS- and PLS-FADS methods and the corresponding TACs obtained by ROI measurements. It can be seen that the PLS-FADS factors agree better with the ROI curves than the factors obtained by the LS-FADS method. Measures D, calculated between the ROI curve and the factor analysis obtained curve, were 0.2874 and 0.1187 for the LS-FADS method and the PLS-FADS method, respectively. We would like to state that the comparison is made to ROI curves which may be biased for the reasons already discussed in the Introduction. However, ROI measurements are widely used for the extraction of the TACs.

TABLE I
COMPARISON OF THE MINIMUM VALUES OF THE MEASURE D WITH CORRESPONDING VALUE OF b OBTAINED BY THE LS- AND PLS-FADS METHODS FOR THE COMPUTER SIMULATIONS

		Noise levels			
		0%	15%	25%	35%
LS-FADS	D	.4413	.4708	.4979	.5155
PLS-FADS	D	.0019	.0312	.0476	.0680
	b	2E3	1E5	2E5	3E5
	$\frac{f_{uni}}{f_{LS}}$	10%	14%	10%	7%

The analysis of the 3-D data (Fig. 5) yields findings similar to those of the 2-D analysis. It is noteworthy that the sixth slice in the 3-D data set is the same slice studied in the 2-D analysis, for which the results are presented in Fig. 4. For the 6th slice in the 3-D data set, FADS with the correction for nonuniqueness gave results that agree better with the ROI measurements than FADS without the correction (Fig. 5) ($D = 0.1953$ for the LS-FADS and $D = 0.0736$ for the PLS-FADS). This is particularly apparent in the tissue curves [Fig. 5(C)]. Also, contrast in the images of the factor coefficients of the LV and the tissue obtained by the PLS-FADS method is improved over the results of the LS-FADS method [Fig. 5(A)–(C)]. For the 3-D LV coefficient image, $\tilde{c}^{\text{LV}}_{tissue}$ changed from 0.63 for the LS-FADS to 0.89 for PLS-FADS method. Also, the value of $\tilde{c}^{\text{LV}}_{\text{RV}}$ was better with the PLS-FADS (1.00) method than with the LS-FADS method (0.87). In the image of tissue coefficients, $\tilde{c}^{\text{tissue}}_{\text{LV}}$ changed from 0.98 to 1.00.

The LS-FADS and PLS-FADS methods were also applied to a patient renal study. Results of the LS-FADS and PLS-FADS analysis, which were applied to the right kidney, are presented in Fig. 6. The top two rows of images present the factor coefficient images for the kidney cortex, background, and pelvis/ureter components obtained using nonunique LS-FADS first row) and PLS-FADS (second row). In the LS-FADS results, the images have similar structures and overlap considerably. This results in factors that do not agree with the ROI curves. These findings are presented in the plots in Fig. 6(A)–(C). In the LS-FADS

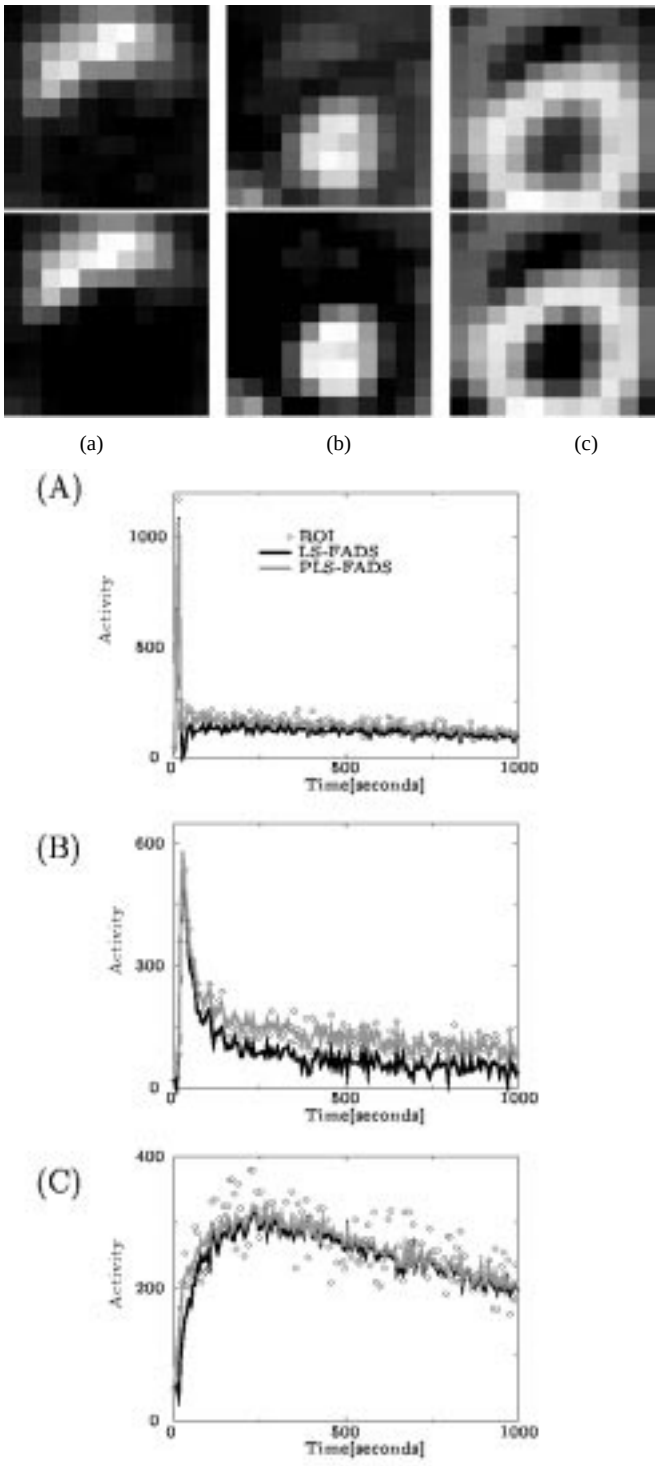

Fig. 4. Results of LS-FADS and PLS-FADS analysis of a ^{99m}Tc-teboroxime canine cardiac study. The top row corresponds to images of factor coefficients for (a) RV blood, (b) LV blood, and (c) tissue components obtained by LS-FADS. The lower row of images was obtained by PLS-FADS. The same gray scale representing values from zero to one is used for each image. Plots (A), (B), and (C) present the factors obtained by the FADS methods with comparison to curves obtained by ROI measurements.

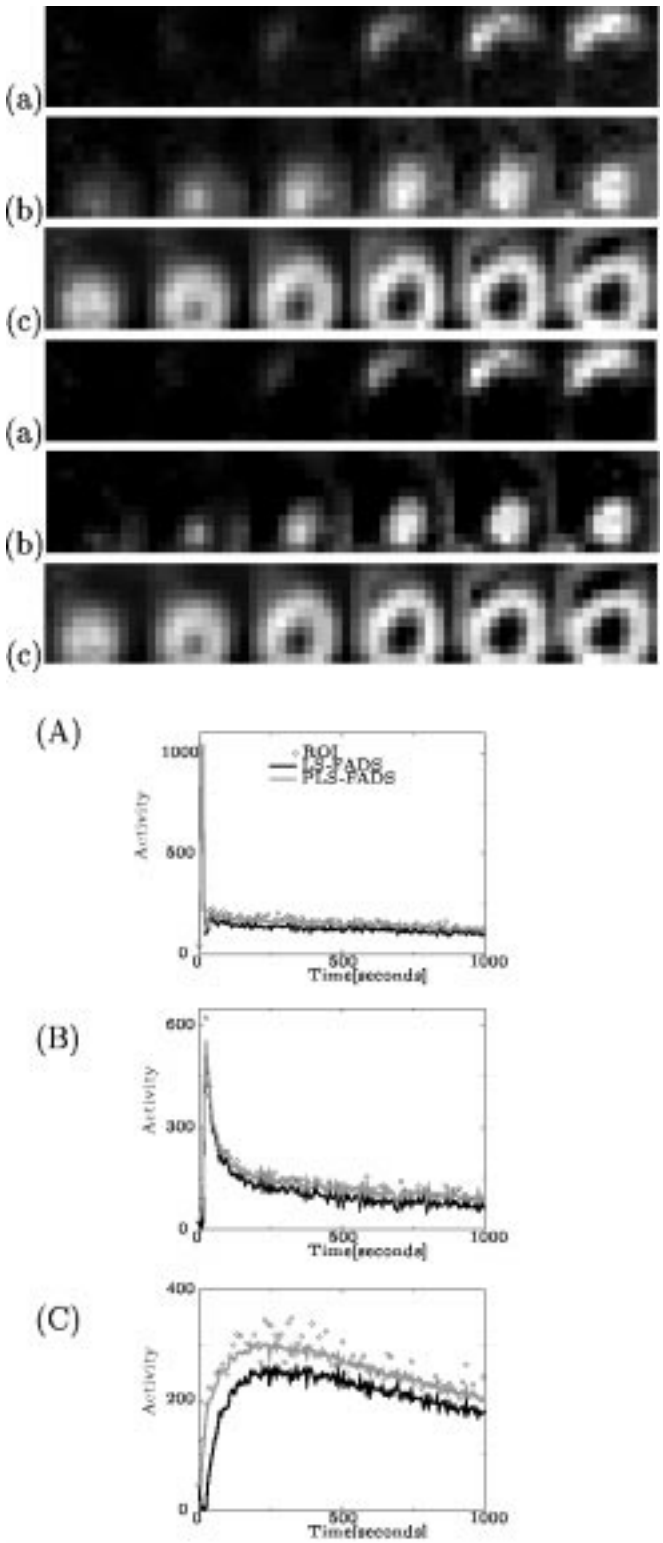

Fig. 5. Results of 3-D LS-FADS and PLS-FADS analysis of a ^{99m}Tc-teboroxime canine cardiac study. The top three rows correspond to images of factor coefficients for 6 consecutive slices for (a) RV blood, (b) LV blood, and (c) tissue components obtained by LS-FADS. The same gray scale representing values from zero to one is used for each image. The lower three rows of images was obtained by 3-D PLS-FADS. Plots (A), (B), and (C) present the factors obtained by the 3-D FADS methods with comparison to curves obtained by ROI measurements.

results (region from 1000 s to 1500 s), the curves appear to be much noisier. This is because the factor coefficient images are similar, which allows the factors to *exchange*, i.e., the factor increase in one curve is compensated for by decreases in the other factors. This is only possible because the images of the factor coefficients are similar and high noise levels are present.

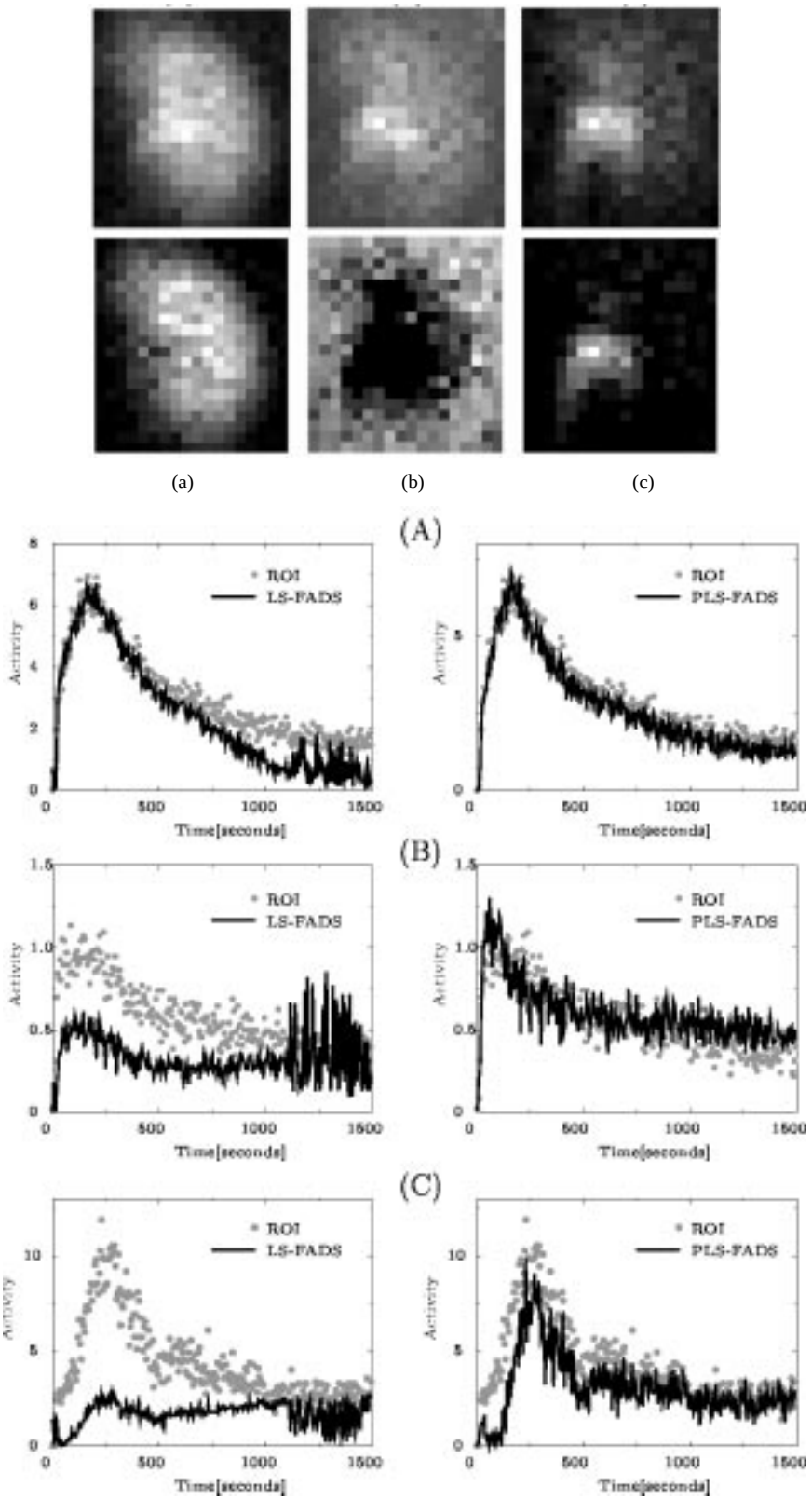

Fig. 6. The results of LS FADS methods of a ^{99m}Tc-MAG$_3$ patient renal study. The top row corresponds to images of factor coefficients obtained by the LS-FADS for (a) kidney cortex, (b) background, and (c) pelvis and ureter components. Second row of images is the same but obtained by the PLS-FADS. The same gray scale representing values from zero to one is used for each image. Plots (A), (B), and (C) present the factors obtained by the PLS-FADS methods with comparison to curves obtained by ROI measurements.

When the PLS-FADS method is applied the obtained curves agree much better with the ROI measurements. The D calculated using the background and the cortex factors between the ROI curves and the FADS obtained curves decreased from $D = 0.2554$ for the LS-FADS to $D = 0.1128$ for the PLS-FADS method. The agreement of background curves is approximate

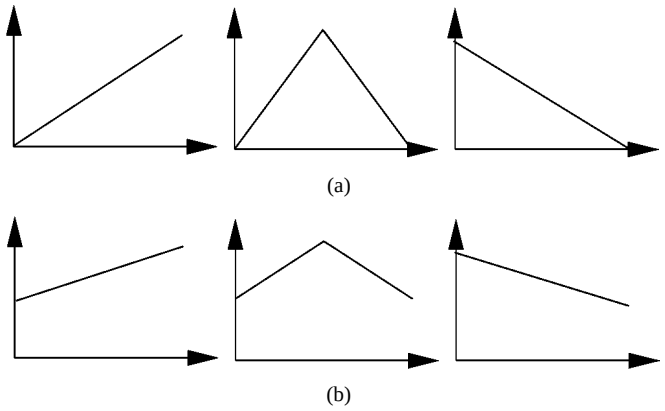

Fig. 7. (a) Factors for which FADS with nonnegativity constraints will give a unique solution and (b) factors used in the computer simulations in this paper.

though because the FADS-obtained background image also contains some of the liver component, which can be seen in the coefficient image of the background as increased activity in the upper right corner of Fig. 6(b), second row. As a result, the corresponding factor is biased by the liver component. The pelvis curves, although similar in shape, differ considerably due to the fact that in the ROI results there is complete overlap of the cortex and pelvis, whereas in the FADS results, these two different physiological regions are separated.

IV. Discussion

The figures presenting the results from computer simulations (Figs. 1 and 2) clearly show that the factor coefficient images are mixed when the FADS method with nonnegativity constraints is employed. For example, in the FADS obtained images of each component, the other components can be seen. Most of the corresponding factors are completely inaccurate and lie far from the simulated curves—this is especially apparent in Fig. 1(A)–(C). The example in Fig. 1 shows the possible severity of nonuniqueness artifacts. This example was specifically chosen to show how inaccurate FADS with nonnegativity constraints can be. On the other hand, it is possible to construct a different computer simulation in which FADS with nonnegativity constraints gives a unique answer. For example, if the factors used are the same as the ones presented in Fig. 7(a), FADS with nonnegativity constraints will give a unique answer. This is because it is impossible to subtract any of those factors from any others without violating the nonnegativity of the factors. This, and the fact that the factor coefficient images also cannot be subtracted from one another without violating nonnegativity, guarantees the uniqueness of the FADS results for this example (for more detail see [7]). Conversely, it was shown in the computer simulations that nonuniqueness artifacts were severe when they were applied to the set of factors presented in Fig. 7(b). It can be concluded that the nonuniqueness effects have a significant impact on the results of FADS, and the severity of the nonuniqueness strongly depends on the study under consideration.

The algorithm presented here introduces unconstrained degrees of freedom due to arbitrary scaling of the factors and factor coefficients. However, this does not affect convergence of gradient-based optimization because directional derivatives of the objective function in directions associated with the scaling ambiguity are zero.

The most problematic issue in the method presented is the selection of the appropriate value of the nonuniqueness penalty parameter b. Fig. 3(a) shows that for the analyzed computer simulation of teboroxime uptake, b needs to be larger than a threshold value in order for the correction to work. The improvement in accuracy of the curve extraction by PLS-FADS is very rapid. It can be seen in the histogram in Fig. 3 that with the same noise level—but different noise realizations, nonuniqueness correction either works (low values of D), or does not work (high values of D), with no values lying in-between. The value of b must also be less than an upper threshold, above which extracted factors and factor coefficients are not accurate because the nonuniqueness term dominates in the objective function and the results of the factor model no longer match the analyzed data. Thus, b has to be in the range between the lower and upper thresholds. As seen in Fig. 3(c), the upper threshold remained the same and the lower threshold changed as noise levels changed. However, the minima of D were always such that the value of $f_{\mathrm{LS}}+f_{\mathrm{neg}}$ was approximately ten times larger than f_{uni}, as can be seen in Table I. This fact was used to select b for the experimental studies. Although the mis-selection of b is a potential problem it is encouraging that (as shown in the computer simulation), the range of "good" b values is wide and, depending on the noise, varies from one to two orders of magnitude. Also encouraging is the fact that when b was selected in the same manner as in the teboroxime study it proved equally useful for the completely different renal study (Fig. 6).

This strategy used for selecting b as described in Section II-B proved to be successful. It worked not only in ^{99m}Tc-teboroxime cardiac imaging and renal imaging as shown in the paper, but also for other dynamic studies not shown here. We found that it worked well for a patient ^{99m}Tc-teboroxime cardiac study with four components (the LV, RV, tissue, and liver), a two-component positron emission tomography liver FDG study, and a dynamic cardiac MRI study.

When using PLS-FADS, the dot product between the factor coefficient images is minimized without violating the nonnegativity constraints, or violating (2), because the constant b in (5) is small. This minimization prevents mixing and creates perfect agreement between the PLS-FADS results and the simulated data [fourth row of Fig. 1 and Fig. 1(A)–(C)]. These effects can also be seen in the experimental data. In the image of the left ventricle in Fig. 4(b) first row, some of the components of the right ventricle and the tissue can be seen. Additional components in this image are removed when PLS-FADS is used [Fig. 4(b) second row].

The same effect can be seen in the images of the tissue components. The tissue image [Fig. 4(c) first row] is biased by the LV and the RV. When PLS-FADS is used the LV and RV contamination is removed from the image of tissue factor coefficients, which increases the contrast in this image [Fig. 4(c) second row] in comparison to the tissue image obtained using the LS-FADS method. Significantly better agreement is achieved between the results of factor analysis and the ROI measurements when the penalized objective function is used. This is especially true for the LV [Fig. 4(B)] and for the RV [Fig. 4(A)] curves.

Different nonuniqueness effects can be seen in the results of FADS for the 3-D data set than can be seen in the FADS re-

sults for the 2-D data set. The tissue curve obtained using the LS-FADS method is much different than the curve obtained by the ROI measurements [Fig. 5(C)]. This disagreement was corrected by applying the nonuniqueness correction (PLS-FADS). A disagreement in the tissue curves obtained by the LS-FADS arises because the LV component completely underlies the myocardial tissue due to the existence of vasculature in the heart muscle. Therefore, the amount of vasculature contained in the tissue curve in the results of the nonunique FADS acquisition is ambiguous. Due to this ambiguity in the 2-D data set the tissue curve obtained by LS-FADS is close to the ROI curve [Fig. 4(C)], and for the 3-D data set it is not [Fig. 5(C)] (see also [7]). Obviously, the tissue ROI curves represent the tissue curve with an addition of a vasculature component. The PLS-FADS removes the disagreement because the nonuniqueness correction minimizes the overlap between the factor coefficients, so that the myocardial tissue and the LV vasculature of the heart muscle are treated as one component, since spatially they occupy the the same space. As a result, the PLS-FADS tissue curve is similar to the one obtained by ROI measurements [Figs. 4(C) and 5(C)]. This can also be seen in the results of the computer simulations (Fig. 2).

Some new nonuniqueness artifacts can be seen in the renal study where the components are exchanged, thereby increasing the noise in the acquired factors [Fig. 6(C)]. This is because images of the factor coefficients are similar. This similarity is removed when the penalty is used in the objective function and the exchange effect is removed in the results of the PLS-FADS. In the renal study, there is a partial overlap between the pelvis component and the cortex. The PLS-FADS method separates these regions [Fig. 6(b) and (c)]. In the factor curve that corresponds to the pelvis, a delay can be seen between the maximum activity in the cortex and the maximum activity in the pelvis. Activity in the pelvis is zero during the first 2 min after injection. These effects cannot be seen on ROI curves because of the overlap between the pelvis and the cortex. Therefore, the ROI curve that corresponds to the pelvis is nonzero from the beginning.

V. Summary

In this paper, we showed the importance of nonuniqueness correction in factor analysis. Based on the LS approach, we developed a simple and straightforward method for correcting the nonuniqueness effects of FADS. The correction was based on minimizing the overlaps between the images of different factor coefficient images. The overlap that occurred was one of the nonuniqueness effects, and, as a result of its minimization, other nonuniqueness effects seen in the factor curves were also minimized. The tradeoff between minimization of the nonuniqueness effects and agreement with the data governed by the value of the parameter b is a limitation of the method. A strategy to overcome this limitation was proposed. The strategy is based on computer simulations and was successfully applied to an experimental canine ^{99m}Tc-teboroxime cardiac study and to a patient ^{99m}Tc-MAG_3 renal study.

Acknowledgment

The authors would like to thank Dr. A. Celler of Vancouver Hospital for providing us with the renal data used in this study. They would also like to thank S. Webb for editing this manuscript.

References

[1] D. C. Barber, "The use of principal components in the quantitative analysis of gamma dynamic studies," *Phys. Med. Biol.*, vol. 25, pp. 283–292, 1980.

[2] R. Di Paola, J. P. Bazin, F. Aubury, A. Aurengo, F. Cavailloles, Y. Herry, and E. Kahn, "Handling of dynamic sequences in nuclear medicine," *IEEE Trans. Nucl. Sci.*, vol. NS-43, pp. 1310–1321, 1982.

[3] V. J. Cunningham and T. Jones, "Spectral analysis of dynamic PET studies," *J. Cereb. Blood Flow Metab.*, vol. 13, pp. 15–23, 1993.

[4] F. O'Sullivan, "Imaging radiotracer model parameters in PET: A mixture analysis approach," *IEEE Trans. Med. Imag.*, vol. 12, pp. 399–412, Sept. 1993.

[5] R. K. Choudhury, "Mixture analysis of multichannel image data," Ph.D. dissertation, Univ. Washington, Seattle, 1998.

[6] A. S. Houston, "The effect of apex-finding errors on factor images obtained from factor analysis and oblique transformation," *Phys. Med. Biol.*, vol. 29, pp. 1109–1116, 1984.

[7] A. Sitek, E. V. R. Di Bella, and G. T. Gullberg, "Factor analysis with *a priori* knowledge—Application in dynamic cardiac SPECT," *Phys. Med. Biol.*, vol. 45, pp. 2619–2638, 2000.

[8] I. Buvat, H. Benali, F. Frouin, J. P. Bazin, and R. Di Paola, "Target apex-seeking in factor analysis on medical sequences," *Phys. Med. Biol.*, vol. 38, pp. 123–128, 1993.

[9] K. S. Nirjan and D. C. Barber, "Factor analysis of dynamic function studies using *a priori* physiological information," *Phys. Med. Biol.*, vol. 31, pp. 1107–1117, 1986.

[10] ——, "The importance of constraints in factor analysis of dynamic studies," in *Information Processing in Medical Imaging*, C. N. de Graaf and M. A. Viergever, Eds. New York: Plenum, 1988, pp. 521–529.

[11] M. Šámal, M. Kárný, H. Surová, E. Maříková, and Z. Dienstbier, "Rotation to simple structure in factor analysis of dynamic radionuclide studies," *Phys. Med. Biol.*, vol. 32, pp. 371–382, 1989.

[12] A. S. Houston, "The use of set theory and cluster analysis to investigate the constraint problem in factor analysis in dynamic structures (FADS)," in *Information Processing in Medical Imaging*, S. L. Bacharach, Ed. Dordrecht, The Netherlands: Nijhoff, 1986, pp. 177–192.

[13] M. Nakamura, Y. Suzuki, and S. Kobayashi, "A method for recovering physiological components from dynamic radionuclide images using the maximum entropy principle: A numerical investigation," *IEEE Trans. Biomed. Eng.*, vol. 36, pp. 906–916, Sept. 1989.

[14] A. Sitek, E. V. R. Di Bella, and G. T. Gullberg, "Factor analysis in dynamic structures in dynamic SPECT using maximum entropy," *IEEE Trans. Nucl. Sci.*, vol. 46, pp. 2227–2232, Dec. 1999.

[15] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C.* Cambridge, U.K.: Cambridge Univ. Press, 1996, pp. 420–425.

[16] B. M. W. Tsui, J. A. Terry, and G. T. Gullberg, "Evaluation of cardiac cone-beam SPECT using observer performance experiments and ROC analysis," *Investigat.. Radiol.*, vol. 28, pp. 1101–1128, 1993.